Problem 1, 50pt

(1) Let the scalar α be defined by $\alpha = \mathbf{y}^\top \mathbf{A}\mathbf{x}$ where $\mathbf{y}$ is $m \times 1$, $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $m \times n$ and $\mathbf{A}$ is independent of $\mathbf{x}$ and $\mathbf{y}$. Represent the derivatives of α with respect to $\mathbf{x}$ and $\mathbf{y}$.

(2) Let the scalar α be defined by $\alpha = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ where $\mathbf{x}$ is $n \times 1$, $\mathbf{A}$ is $n \times n$ and $\mathbf{A}$ is independent of $\mathbf{x}$. Represent the derivatives of α with respect to $\mathbf{x}$.

(3-5) Represent the derivative of the following scalar functions with respect to $\mathbf{X} \in \mathbb{R}^{D \times D}$.

(3) $f(\mathbf{X}) = tr(\mathbf{X}^2)$. Here, $tr(A)$ is the trace of a square matrix A.

(4) $g(\mathbf{X}) = tr(\mathbf{X}^3)$.

(5) $h(\mathbf{X}) = tr(\mathbf{X}^k)$ for $k \in \mathbb{N}$.

Problem 2, 20pt

For a given data set $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^N$, derive the closed-form solution for $\mathbf{w}$ that maximizes the following probability $\boldsymbol{P}$.

$$\boldsymbol{P} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - f(\mathbf{x}_i; \mathbf{w}))^2 \right), \tag{1}$$

with $f(\mathbf{x}; \mathbf{w}) = \mathbf{w}^\top \mathbf{x}$ for $\mathbf{x}, \mathbf{w} \in \mathbb{R}^D$.

Problem 3, 20pt

The Frobenius dot product $\langle \mathbf{A}, \mathbf{B} \rangle$ is defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle = tr(\mathbf{A}^\top \mathbf{B}), \tag{2}$$

for $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times N}$ using the scalar function $tr(\cdot)$ for the trace of a matrix,

$$tr(\mathbf{M}) = \sum_i^N M_{ii}, \quad \text{for } \mathbf{M} \in \mathbb{R}^{N \times N}. \tag{3}$$

Find the derivative of $\langle \mathbf{A}, \mathbf{B} \rangle$ with respect to $\mathbf{A}$ using the definition of the matrix derivative:

$$\left[\frac{d}{d\mathbf{A}} \langle \mathbf{A}, \mathbf{B} \rangle \right]_{ij} = \frac{\partial}{\partial A_{ij}} \langle \mathbf{A}, \mathbf{B} \rangle. \tag{4}$$

Problem 4, 20pt

We are given a scalar function $f(\mathbf{X}) = \mathbf{w}^\top \mathbf{X} \mathbf{w}$ for $\mathbf{w} \in \mathbb{R}^D, \mathbf{X} \in \mathbb{R}^{D \times D}$. Find the derivative of $f(\mathbf{X})$ with respect to the vector $\mathbf{w}$. Use the definition of the vector derivative $\left[\frac{df}{d\mathbf{w}}\right]_k = \frac{\partial f}{\partial w_k}$, where w_k is the k-th element of $\mathbf{w}$.

Problem 5, 40pt

Assuming the following images are given: Each image contains two numbers (0 $\sim$ 9 for each number) and is in black and white. The size of the image is 20x20. We want to construct a machine-learning algorithm to predict the two numbers contained in the given image. For example, in the below image, the goal is to predict 10, 48, 59, 62, and 73 as outputs. Define the input and output of your machine-learning algorithm. Also, define the objective function, propose a structure for learning, and explain the reasons for it. (Provide any assumptions necessary to solve the problem.)